Supplementary Information

Deep Learning Based Accurate Hepatic Steatosis Quantification for Histological Assessment of Liver Biopsies

Mousumi Roy, Fusheng Wang, Hoang Vo, Dejun Teng, George Teodoro, Alton B. Farris, Eduardo Castillo Lion, Miriam B. Vos, and Jun Kong

DELINEATE Architecture

Our input training dataset is denoted as $D = \{(X_i, Y_i), i = 1, 2, ..., N\}$, where N is the total number of image patches. Each input image patch X_i has 512×512 pixels in size. The corresponding label for X_i is a steatosis mask Y_i. As this is a two-class instance segmentation problem, we use label 1 and 0 to represent steatosis region and background, respectively. The segmentation process has two steps. First, regions and boundaries of steatosis droplets are predicted independently. (X_i, Y_i^r) and (X_i, Y_i^b) are paired inputs for region extraction and boundary prediction models, respectively. The ground truth region masks $\{Y_i^r\}$ are generated from the ground truth boundary masks $\{Y_i^b\}$. The combined output from steatosis region and boundary models (X_i^c, Y_i^c) in stage one is the input to FCN in the second stage for further instance segmentation.

The region extraction model aims at identifying steatosis components from the image background. Inspired by the state-of-the-art solution from the FCN family model U-Net [1], we design our region extraction model by including multiple encoding and decoding layers. U-Net is specifically designed for bio-medical image segmentation tasks where each pixel is assigned a class label from a pre-specified class set. Different levels of contextual feature maps are extracted by the encoding or down-sampling layers for feature extraction. The decoding or up-sampling layers act as a pixel-wise predictor that generates the desired probability masks along with feature maps from the down-sampling path.

Although discriminating high-level features are extracted in the encoding layers, the detailed spatial information can be lost due to feature map resolution reduction by max-pooling and strided convolution in the down-sampling path. In the U-Net model, this information loss is compensated by the aggregated high resolution feature maps from the encoder layers to the corresponding decoder layers. For further reduction

in information loss, dilated convolution specifically designed for dense prediction has been proposed for convolutional network module [2]. Dilated convolution enriches the network with an exponential expansion of the receptive field without down-sampling or resolution loss. Replacing the down-sampling operation with dilated convolutions results in a semantic segmentation system with less space-invariance and better accuracy. This is demonstrated by a prior research where atrous convolution along with atrous spatial pyramid pooling (ASPP) was shown as a powerful tool for dense object segmentation at multiple scales [3].

Our steatosis region extraction model consists of four down- and four up-sampling layers, respectively. We modify the U-net model by stacking dilated convolution layers (depth = 3) at the bottleneck block where the feature maps have the lowest resolution. This architecture, named as dil-Unet, is illustrated in Fig.1. The dilated convolutional layers in U-Net improve the steatosis region recognition as demonstrated in the result section. We use 512×512 input images in our experiment. The kernels used in each convolutional layer are initialized by the standard Xavier initialization [4] and the bias is initialized with zero. Xavier initialization or normalized initialization can be described as follows:

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \tag{1}$$

where W represents the initialized weights and n_j is the size of the j^{th} convolutional layer [5].

We train our region extraction network to minimize the softmax cross entropy loss L^r between the prediction map P^r and the target Y^r:

$$L^r = -\sum_{c=1}^{K} \sum_{i,j \in \Omega} y^r(i,j) \log(P^r(i,j|c,w^r)) \tag{2}$$

where $K = 2$ is the number of classes, i.e. steatosis and background; Given the trained network parameter set $\{w^r\}$, y^r is the binary indicator of the true label at pixel (i,j) and $P^r(i,j|c,w^r,b^r)$ is the output of soft-max activation layer indicating the probability of the pixel (i,j) having label c. The softmax output from the final layer of region model is:

$$P^r(i,j|c,w^r) = \frac{\exp(a_c(i,j|w^r))}{\sum_{k=1}^{K} \exp(a_k(i,j|w^r))} \tag{3}$$

where $a_k(i,j|w^r)$ represents the activation in feature channel k at the pixel (i,j). We use Adam optimizer [6] along with exponential learning rate decay to optimize the parameter set $\{w^r\}$ by back-propagation.

Although dil-Unet based segmentation is insufficient to divide overlapped steatosis droplets by itself in our experiments, the probability masks produced by dil-Unet provide supporting information for further analysis. As a result, we further develop a steatosis boundary detection model to delineate the hidden boundaries of the overlapped steatosis droplets.

Recent Deep Neural Networks (DNN) based studies on nuclei segmentation and gland instance segmentation have suggested that combination of region, location and edge information can achieve state-of-the-art results [5, 8, 9, 7]. With our data, we notice that overlapping steatosis droplets remain connected in the predicted region mask. Joint use of steatosis boundary information can compensate for the loss of fine-resolution spatial information due to the max-pooling and strided convolution in the encoder-decoder model for region detection. Thus, we integrate the boundary information with the region prediction results.

As the number of boundary pixels is much less than that of background pixels, this imbalance leads to convergence issue when the dil-UNet model is used for the boundary detection. Multiple DNN models, including N4-Fields [10], DeepContour [11], DeepEdge [12], CSCNN [13], focus on automatic hierarchical feature learning and are used for object boundary detection in images of natural scenes. In our study, we use the Holistically-Nested Network (HNN) [14], to learn the steatosis boundary maps. Similar to the previous work [14], the CNN architecture derived from VGGNet model is adopted in our boundary detection module. Initializing the weights with the pre-trained network [14], we further train the HNN model with our training data (X, Y^b). HNN consists of five convolutional stages having stride values 1, 2, 4, 8, and 16 respectively and distinct receptive field sizes [14]. Additionally, a HNN model has M side-output layers serving as classifiers with weights $w = (w^{(1)}, ..., w^{(M)})$. We use W to represent the complete set of parameters for standard network layers. The resulting objective function can be represented as:

$$L_1(W, w) = \sum_{m=1}^{M} \alpha_m l^m(W, w^m) \tag{4}$$

where l^m is the image level loss function from side output m and can be computed over all pixels of training image pair (X, Y^b). Each side output l^m is refined for minimization over iterations.

In this model, the sigmoid activation function $\sigma(\cdot)$ is used to compute the class probability at each image pixel (i, j):

$$P^b(i, j|X; W, w^m) = \sigma(a^m(i, j)) \tag{5}$$

Steatosis edge map prediction at the side output layer m is $\hat{Y}_1^{(m)}(i, j) = \sigma(a^m(i, j))$ where $a^m(i, j)$ is the

activation value from the side output layer m.

Finally, we add to the network a "weighted-fusion" layer that is trained in parallel during the training phase [14, 15]. The fusion layer loss function is defined as:

$$L_2(W, w, h) = Dist(Y, \hat{Y}_2) \tag{6}$$

where $\hat{Y}_2 = \sigma(\sum_{m=1}^{M} h_m a^m(i,j))$ has fusion weights $\{h_i\}$. $Dist(\cdot)$ is the cross-entropy loss with the fused predictions and the ground truth label maps. Optimal parameters are found by objective function minimization with stochastic gradient descent and back propagation in training:

$$(W, w, h)^* = \arg\min(L^b) \tag{7}$$

where $L^b = L_1(W, w) + L_2(W, w, h)$.

In the testing phase, the prediction is generated from both side output layers and the weighted-fusion layer:

$$(\hat{Y}_1^{(1)}, \cdots, \hat{Y}_1^{(M)}, \hat{Y}_2) = HNN(X, (W, w, h)^*) \tag{8}$$

It can be insufficient to use the average of all outputs from equation 8 as the boundary detection result [14]. Instead, in our study, we use the 5^{th} side output to represent steatosis boundaries, i.e. $\hat{Y}^b = \hat{Y}_1^{(5)}$, as it presents clean results by visual inspections.

After intensive experiments, we noticed that neither region nor boundary information by its own is sufficient for accurate steatosis droplet segmentation. In addition, neither direct combination nor simple concatenation of the two channels of outputs provides precise boundary information for overlapped steatosis droplets. False positive and missing segmentation results were also noticed in region and boundary outcomes as shown in Fig. 2. These results require an effective mechanism to integrate region and boundary information for enhanced segmentation performance. Unlike classical segmentation methods, deep learning based frameworks, once trained, can be used for prediction without further explicit parameter change. Therefore, we use a deep learning model in the integration phase to make the whole framework more generic without explicit parameter-tuning. In our implementation, we use Fully Convolutional Network (FCN) [16], a simple and fast segmentation model presented in Fig.1, as our integration network to generate the final prediction result. Specifically, we use FCN-8s with skip connections from pool3 and pool4 for better deep semantic information

integration on the down-sampling path. The final output has three channels representing the probabilities of each pixel being background, boundary, or region class, respectively. Weight vectors for this network are denoted as w^c. The outputs of region and boundary modules $(\hat{Y}^r, \hat{Y}^b)$ are combined by the FCN model which generates a final 3-class prediction result $P^c(i, j|\hat{Y}^r, \hat{Y}^b; w^c)$. We train the integrated network with softmax cross entropy loss and Adam optimizer [6], with dropout probability 0.3 to overcome the over-fitting problem.

Results

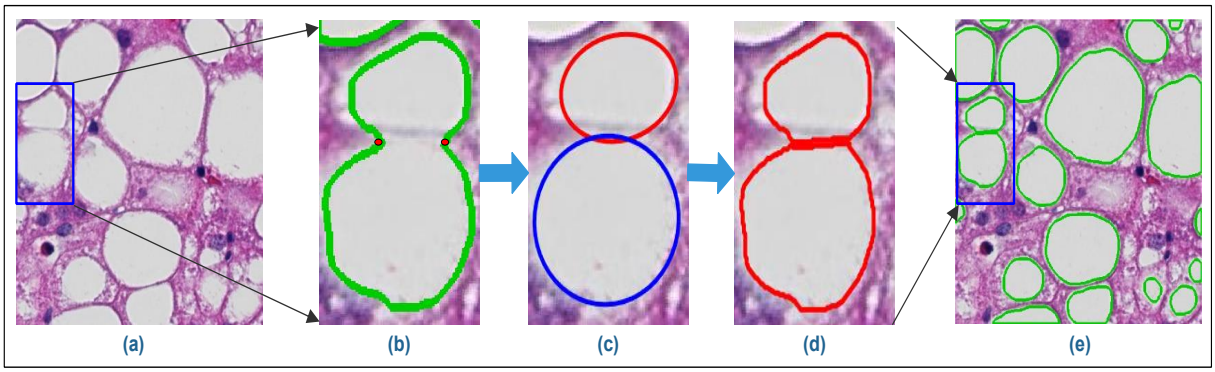


Figure 1: In the post-processing stage, clumped steatosis droplets are segregated by an ellipse fitting quality assessment method [17]: (a) Overlapped steatosis instances; (b) Curvature points are detected as red dots; (c) Identified ellipse representations for steatosis boundaries; (d)Recovered steatosis boundaries; and (e) Detected steatosis droplets with finalized separating borders

Fig. 1 illustrates all steps in the post-processing described in the main manuscript. Fig. 2 demonstrates typical segmentation results where large areas marked by "1" are detected as steatosis regions by mistake in the first stage and filtered in the final prediction result (marked by "2") using the integration model FCN-8s in the second stage. The exclusion of such falsely detected regions by the third integration network improves the overall performance of our DELINEATE model. The qualitative comparison results across our proposed method, FCN, DeepLab and variations of our method are presented in Fig. 3 where the recovered steatosis boundaries in green are overlaid with the original images. Additionally, we plot Receiver Operating Characteristic curves for all comparison methods in Fig. 4. We further demonstrate quantitative steatosis segmentation accuracy profiles for one whole tissue component in Fig. 5 and Fig. 6. Fig. 5 presents how performance measures are spatially changed over tissue patches, while Fig. 6 illustrates such variation in steatosis components.

Table 1: Summary of the key clinical information of the analyzed patient dataset

					NAFLD (n=36)
Parameter	Age, mean (SD)				14.9 (2.59)
	Gender			Female	12 (33.3%)
				Male	24 (66.7%)
	BMI z-score, mean (SD)				2.30 (0.52)
	BMI Category			Normal	1(2.8)
				Overweight	4(11.1)
				Obese	31(86.1)
	NAFLD Diagnosis			NAFL	28 (77.8%)
				NASH	8 (22.2%)
	Steatosis Assessment	Liver Biopsy	Steatosis	Grade 0 (Less than 5%)	4 (11.1%)
				Grade 1 (Between 5-33%)	9 (27.8%)
				Grade 2 (Between 34-66%)	10 (22.2%)
				Grade 3 (More than 66%)	13 (38.9%)
			Total Steatosis, mean (SD)		46.3 (29.9)
			Macrovesicular Steatosis, mean (SD)		44.5 (28.7)
		MRI	Volumen Fraction of Fat %, mean (SD)		11.1 (8.20)
Histology	Lobular Inflammation			None	8 (22.2%)
				Less than 2/20 x per HPF	27 (75.0%)
				Between 2-4/20x per HPF	1 (2.8%)
	Ballooning			None	25 (69.4%)
				Few	11 (30.6%)
	Portal Inflammation			None	33 (91.7%)
				Mild	3 (8.3%)
	Fibrosis			0: None	13 (36.1%)
				1A: Zone 3, perisinusoidal, delicate	15 (41.7%)
				1B: Zone 3, perisinusoidal, dense	3 (8.3%)
				1C: Portal, periportal only	1 (2.8%)
				2: Zone 3 perisinusoidal and portal/periportal	3 (8.3%)
				3: Bridging Fibrosis	1 (2.8%)

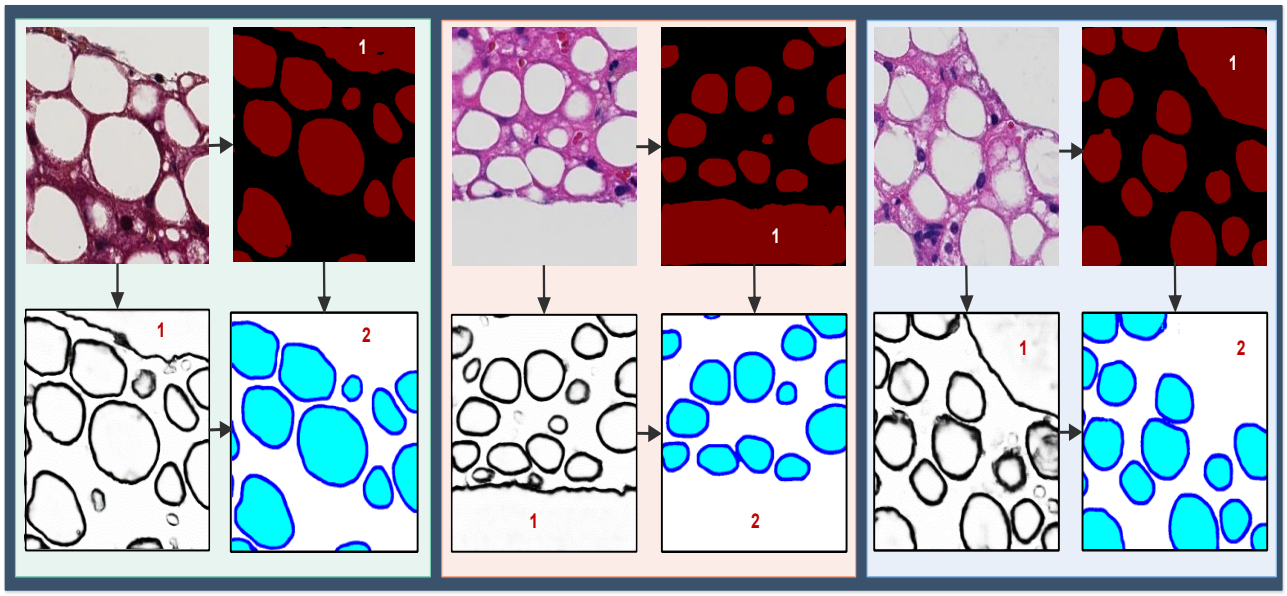


Figure 2: Three typical regions with improved steatosis segmentation by the region and boundary integration network are presented. Top-Left: input image; Top-Right: output from the region extraction model; Bottom-Left: output from the boundary detection model; and Bottom-Right: final output of the integration model. False steatosis region captured by the region prediction model and corrected false steatosis regions by the final integration network are labeled by "1" and "2" in each example region.

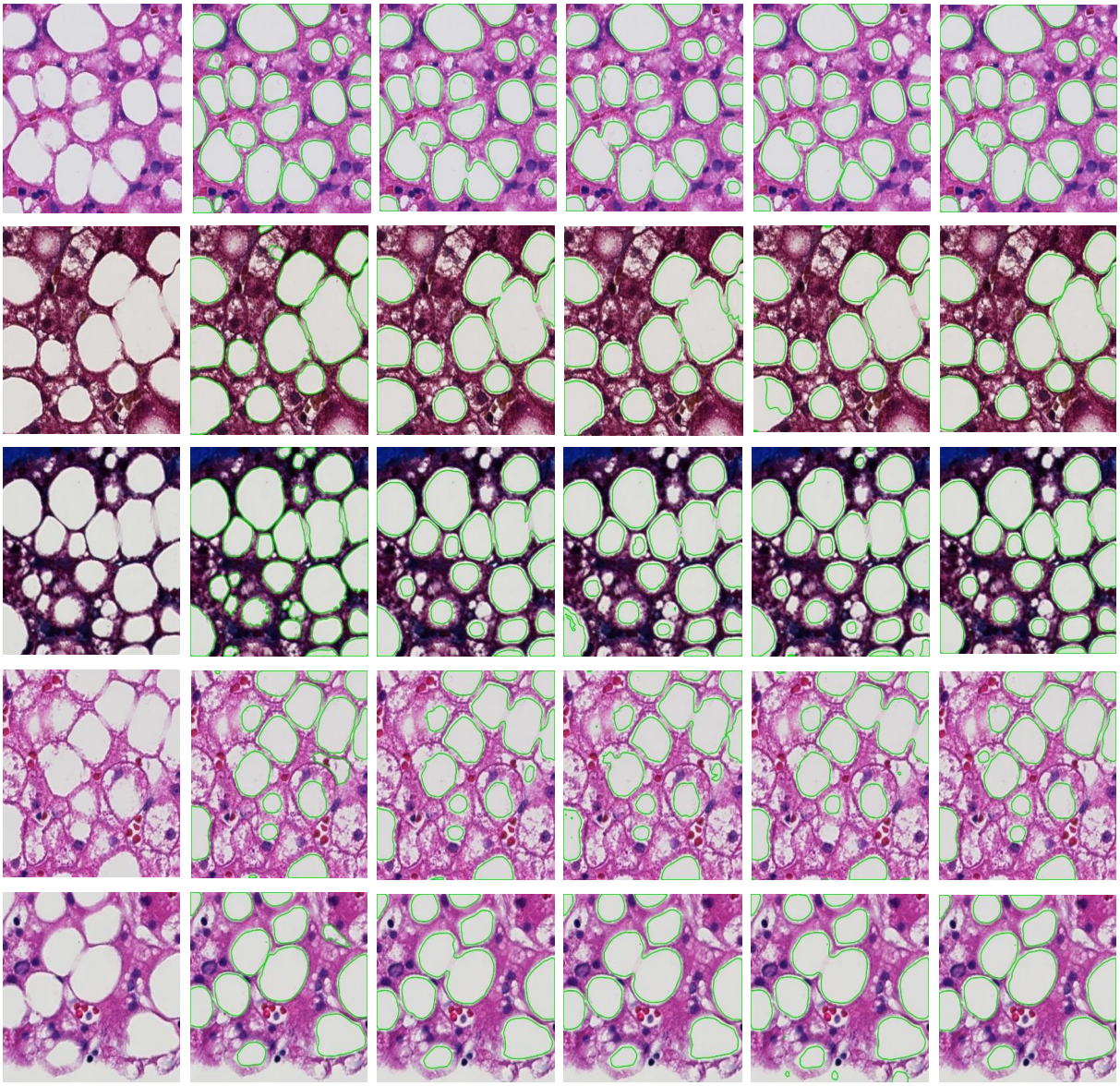


Figure 3: Automatically recognized steatosis droplet boundaries in green are overlaid with representative image patches (in rows). From left to right: the original image, ground truth, results from FCN [16], DeepLab V2 [3], `Unet+Unet+Unet` (one variation of our proposed model), and `dil-Unet+HNN+FCN-8s` (our proposed DELINEATE model). These images present overlapped steatosis droplets with large variations in color and shape. Note that our proposed DELINEATE model can better delineate overlapped steatosis boundaries than other models in the comparison set.

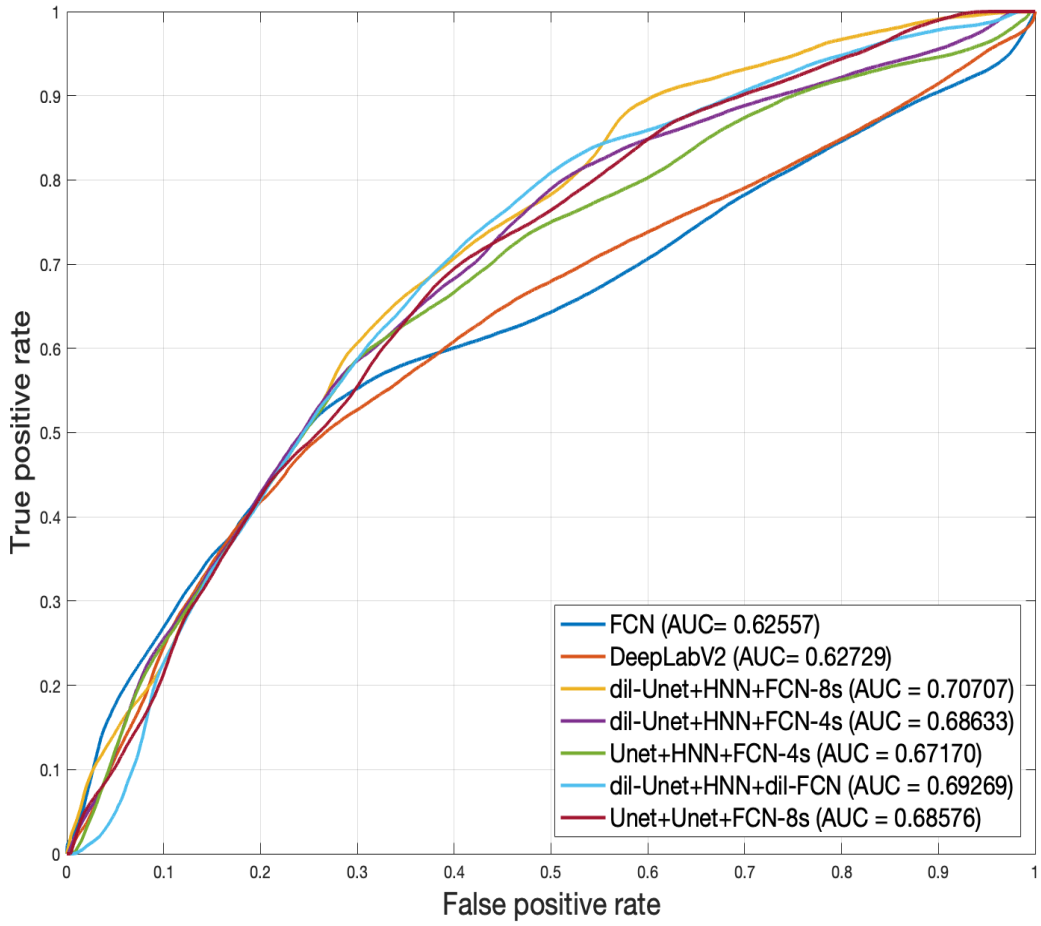
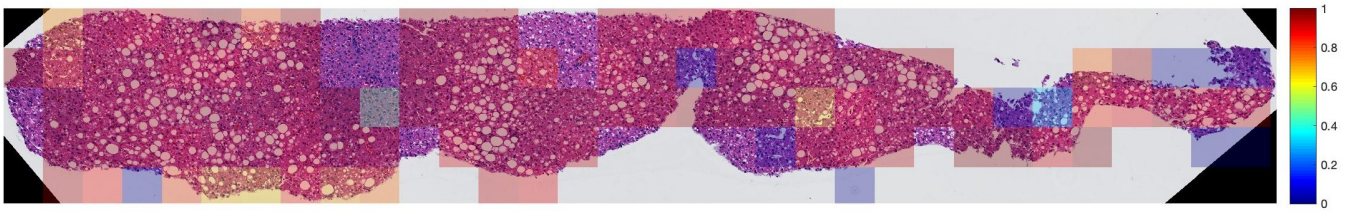
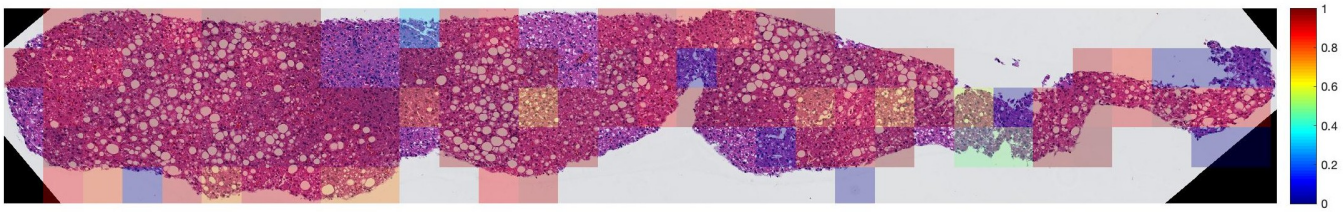


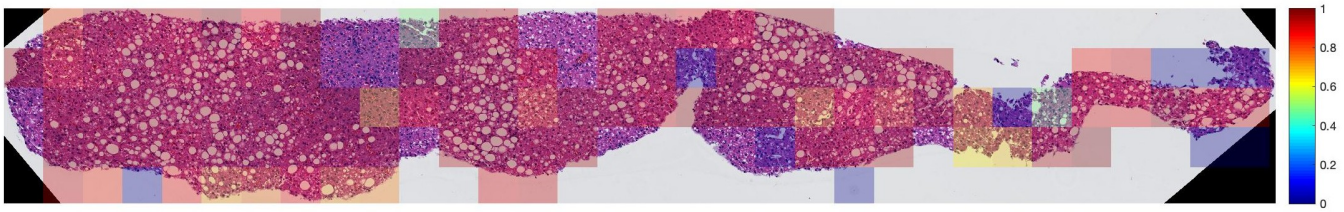
Figure 4: Receiver Operating Characteristic (ROC) plots of multiple methods for comparison.



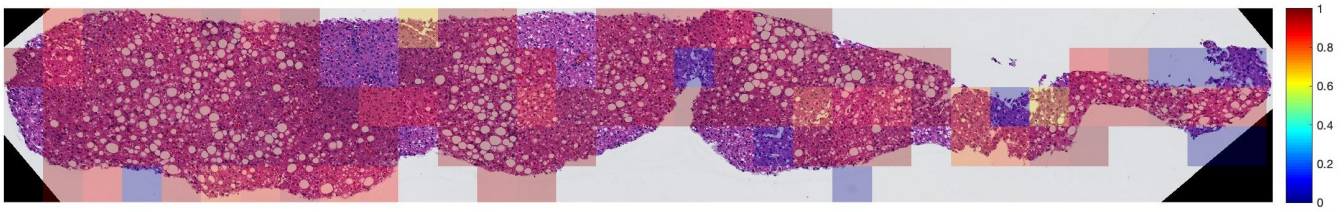
(a)



(b)

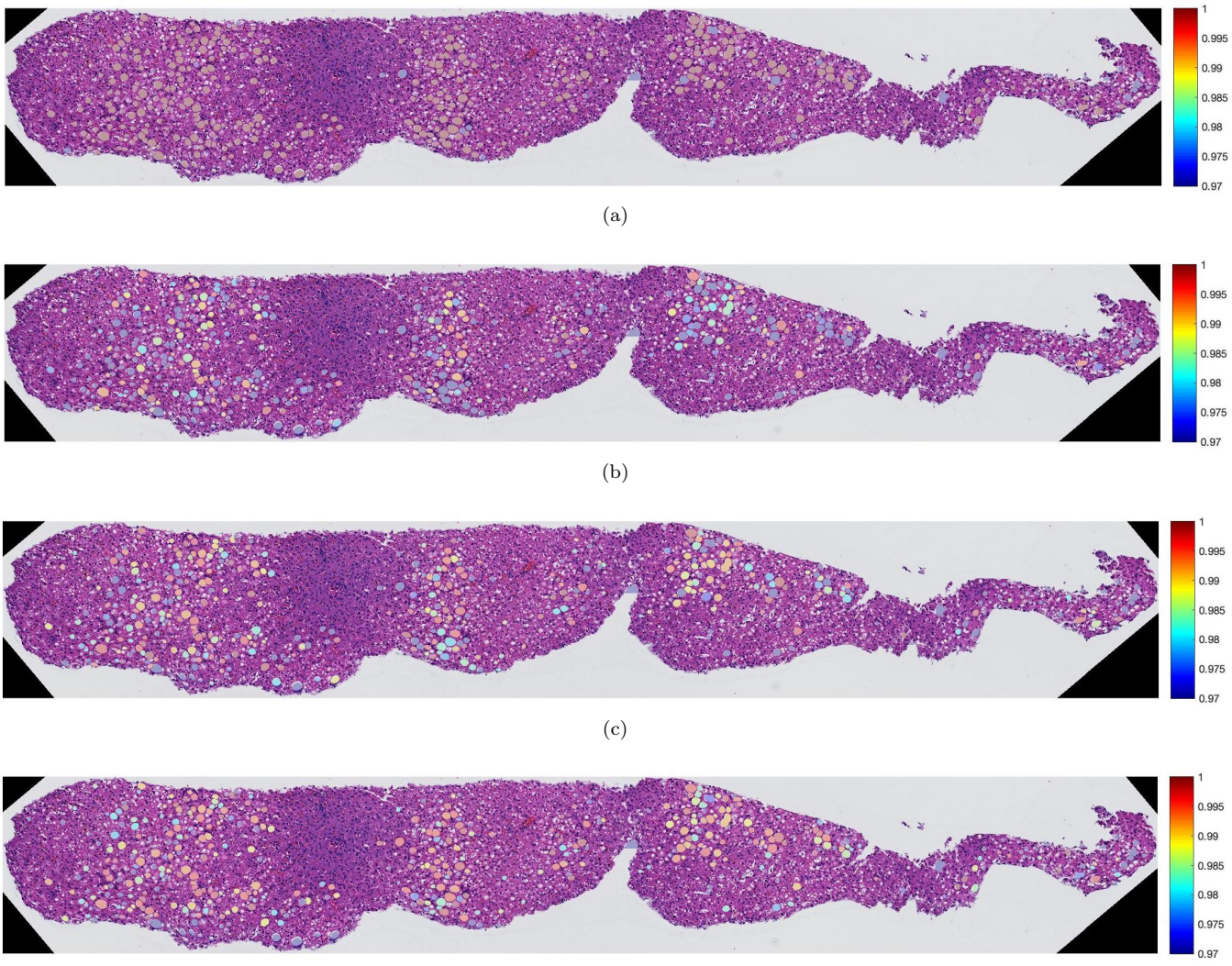


(c)



(d)

Figure 5: Heat maps of image patch-wise performance variations of the proposed DELINEATE model, as measured by (a)Precision, (b)Recall, (c)F1-Score, and (d) object-wise Dice index, are superimposed on an original whole-slide liver tissue component image.



(a)

(b)

(c)

(d)

Figure 6: Performance variations of our proposed DELINEATE model are illustrated by color coded steatosis regions in a whole-slide liver tissue component image. Heat map plots of (a)Precision, (b)Recall, (c)F1-Score, and (d) object-wise Dice index value for each steatsosis droplet are superimposed on the original tissue image.

Dataset

The original clinical datatset information is in the file dataset.xlsx available at GitHub [18]. The key clinical summary information is presented in Table 1.

References

[1] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.

[2] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions. International Conference on Learning Representations (ICLR). 2016. https://arxiv.org/abs/1511.07122

[3] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017;40(4):834-48.

[4] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics. 2010:249-56.

[5] Cui Y, Zhang G, Liu Z, Xiong Z, Hu J. A deep learning algorithm for one-step contour aware nuclei segmentation of histopathology images. Medical & biological engineering & computing. 2019;57(9):2027-43.

[6] Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. 3rd International Conference for Learning Representations. 2015. http://arxiv.org/abs/1412.6980

[7] Chen H, Qi X, Yu L, Dou Q, Qin J, Heng P-A. DCAN: Deep contour-aware networks for object instance segmentation from histology images. Medical image analysis. 2017;36:135-46.

[8] Xu Y, Li Y, Wang Y, Liu M, Fan Y, Lai M, et al. Gland instance segmentation using deep multichannel neural networks. IEEE Transactions on Biomedical Engineering. 2017;64(12):2901-12.

[9] Khoshdeli M, Parvin B. Deep Learning Models Delineates Multiple Nuclear Phenotypes in H&E Stained Histology Sections. ArXiv. 2018;abs/1802.04427. https://arxiv.org/abs/1802.04427

[10] Ganin, Y., & Lempitsky, V. (2014, November). N^4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms. In Asian Conference on Computer Vision (pp. 536-551). Springer, Cham.

[11] Shen, W., Wang, X., Wang, Y., Bai, X., & Zhang, Z. (2015). Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3982-3991).

[12] Bertasius, G., Shi, J., & Torresani, L. (2015). Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4380-4389).

[13] Hwang, J. J., & Liu, T. L. (2015). Pixel-wise deep learning for contour detection. arXiv preprint arXiv:1504.01989. https://arxiv.org/abs/1504.01989

[14] Xie S, Tu Z. Holistically-nested edge detection. In Proceedings of the IEEE international conference on computer vision. 2015:1395-403.

[15] Roth HR, Lu L, Farag A, Sohn A, Summers RM. Spatial aggregation of holistically-nested networks for automated pancreas segmentation. In International conference on medical image computing and computer-assisted intervention: Springer, Cham; 2016. p. 451-9.

[16] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:3431-40.

[17] Roy M, Wang F, Teodoro G, Vos MB, Farris AB, Kong J. Segmentation of overlapped steatosis in whole-slide liver histopathology microscopy images. In 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC): IEEE; 2018. p. 810-3.

[18] Roy M, Wang F, Vo H, Teodoro G, Farris AB, Lion EC, et al. Data from "DELINEATE" Github. 2020. Available from: https://github.com/StonyBrookDB/DELINEATE